

# Εργασία 2

---

## Εισαγωγή

Η εργασία αυτή έχει ως στόχο την εμπέδωση των αντικειμένων της μάθησης από δεδομένα πολλαπλών ετικετών και της μάθησης από σάκους περιπτώσεων. Στην εργασία θα δουλέψετε με το σύνολο δεδομένων [DeliciousMIL](#), το οποίο περιέχει ένα σύνολο εγγράφων, όπου το κάθε έγγραφο θεωρείται ως ένας σάκος από προτάσεις. Μπορείτε να τα διανυσματοποιήσετε με το module [feature\\_extraction.text](#) της sklearn.

## Μέρος Α

Στο πρώτο μέρος της εργασίας θα δουλέψετε στο αντικείμενο της μάθησης από δεδομένα πολλαπλών ετικετών. Εφαρμόστε δύο τεχνικές μάθησης από δεδομένα πολλαπλών ετικετών στα παραπάνω δεδομένα, όπου η αναπαράσταση κάθε εγγράφου αφορά στις λέξεις που περιλαμβάνει (bag of words). Παρουσιάστε και σχολιάστε τα αποτελέσματα των πειραμάτων σας με βάση τα δοθέντα train και test splits. Tip: σε τέτοιου τύπου αραιές αναπαραστάσεις, καλή επίδοση έχουν συνήθως τα γραμμικά μοντέλα ή/και τα γραμμικά SVM.

## Μέρος Β

Στο δεύτερο μέρος της εργασίας θα δουλέψετε στο πρόβλημα της μάθησης από σάκους περιπτώσεων. Εστιάστε στην συχνότερη από τις 20 κλάσεις, ώστε να αντιμετωπίσετε ένα δυαδικό πρόβλημα ταξινόμησης. Θεωρείστε ότι κάθε έγγραφο είναι ένας σάκος από προτάσεις. Ομαδοποιήστε τις προτάσεις του συνόλου εκπαίδευσης και αναπαραστήστε κάθε έγγραφο με βάση τις ομάδες στις οποίες ανήκουν οι προτάσεις του. Δοκιμάστε επιπλέον ένα μοντέλο που αναπαριστά κάθε έγγραφο με βάση όλες τις προτάσεις του. Παρουσιάστε και σχολιάστε τα αποτελέσματα των δύο αυτών προσεγγίσεων.

## Λογιστικά

Οι εργασίες μπορούν να γίνουν είτε ατομικά, είτε σε ομάδες 2 ατόμων. Θα πρέπει να παραδώσετε από ένα Colab notebook για κάθε μέρος της εργασίας. Κάθε notebook θα περιέχει τον κώδικα που χρησιμοποιήσατε, θα παρουσιάζει τα αποτελέσματα των πειραμάτων σας και θα τα σχολιάζει. Θα πρέπει να παραδώσετε μέσω του elearning ένα αρχείο zip με τα δυο notebooks, αφού τα κατεβάσετε από το Colab ως ipynb αρχεία. Προθεσμία υποβολής: 21/5/2023. Για απορίες, χρησιμοποιήστε το φόρουμ «Questions on the Assignments» στο elearning.